

Research on Natural Interactive 3D Registration Algorithm Based on the Detection of Human Hand Features

J.C. Fan

School of Information and Communication
North University of China
Shanxi, China

Y.Z. Liu

School of Automation
Beijing University of Post and Telecommunications
Beijing, China

Abstract—Aiming at hand feature detection and recognition problem in the natural interaction, especially for the problem that finger was ignored or undetected in the current method, we propose, an improved Adaptive Mean-SHIFT(CamShift) algorithm which combined skin color detection to track and detect hand information accurately. In terms of in the natural interaction, the hand's position and posture could be acquired effectively when it's moving and flipping. We proposed a hand feature point extracting method that based on improving corner feature curvature detection method, then divided these feature points into different group effectively, using weighted algorithm to describe these feature point. This method improve the accuracy of detection and ensure the completion of 3D registration. Experimental result shows than this algorithm detect more accurate than traditional mean-shift algorithm, has stronger robustness for the target occlusion and posture change video sequence.

Keywords—hand feature; tracking and detection; 3D registration; adaptive; mean-shift(CamShift); natural interaction

I. INTRODUCTION

With the development of computer vision technology, more and more attention has been paid on natural interaction, the detection and location of hand in the virtual-real interaction system play a decisive role in subsequent calibration, operation and virtual-real fusion. This machine vision-based virtual-real interaction using a natural feature registration method, not only reduces the cost of the equipment and improves the user's interactive experience, but also achieves the natural interaction. However, the natural change of hand shape has the high dimension features, well as the unpredictable occlusion between the fingers, making it very difficult to extract the hand features and the real-time reconstruction of the detailed hand structure based on the feature extraction. Therefore, the vision-based gesture recognition has been a challenging interdisciplinary research topic [1]. Addressing the augmented reality system, this paper proposed an optimal combination of the adaptive mean shift algorithm combined the skin color detection to achieve precise position of the hand, using the improved curvature detection method to extract the feature points, obtaining the description of the hand features via K-means clustering so as to provide the precise position/pose information for subsequent 3D registration. Comparing with traditional methods, the method

has solved the problems such as the higher environmental conditions and the limited detection ability of the original method, further improved the speed and precision of the hand tracking. The novel method is able to maintain relatively accurate tracking in complex environment and shows a better robustness. It has improved the precision of the description of feature points, and obtained higher accuracy of the hand in the operations such as rotation and translation.

II. RELATED WORKS

In recent years, many scholars have conducted the in-depth study of the issues of the hand virtual-real interaction in the natural interaction system; many outstanding achievements have been made in the gesture interaction method based on a single camera. In literature [2], the specific hand gesture as an interactive information carrier, does not require a direct interaction between the human hand and virtual object, but this kind of methods has the implied in-depth ambiguities. Literature [3] has put forward the grasping recognition method based on visual perception, which enhances the realism of the user in the interactive operation through his hand correctly grasping the virtual objects. But this method is limited to grasp the cylindrical object. In literature [4] the hand is regarded as a whole in the interaction, ignoring dynamic details of the hand, however, the details of the hand shall be distinguished in the fine interactive application, how to effectively extract the detail features and how to effectively obtain the interactive information are the key points. In the literature [5] the fingertip features are obtained through calculating the K curvature of its contour, and taking the extreme according to the K curvature of the point of the fingertip. The literature [1] presents a human hand natural interactive method based on multiple view geometry, which can realize the interactive operation between human and virtual object without any mark, using a multi view based sparse stereo 3D structure reconstruction method to eliminate the hidden in-depth ambiguities, but it needs to set 4 cameras, and the problem of the non color object cannot be segmented in dynamic perspective exists. The literature [6] uses the hidden Markov model to realize the hand detection under the complex background, which has high detection accuracy, but the detection speed is low. However, with the development of computer hardware, it is believed that this method will become one of the most important hand detection methods.

III. FEATURE POINTS EXTRACTION AND DESCRIPTION

The natural features of hand in the natural interactive registration system is composed of two parts: 1) the hand recognition detection 2) feature extraction and description, Hand feature extraction is the main part of the realization of 3D registration[7]. The CamShift algorithm can accomplish the real-time and adaptive tracking of the target, but the target detection results are presented through a rotating elliptical shape, which is not conducive to the follow-up study and treatment, most parts of the hand can be covered by the elliptical shape, but the part of the fingertips lies at the outside [8] of the hand. In the tracking process, it is difficult to be covered by the elliptical area, however the fingertips as prominent area of points of the hand features, which is very important for the extraction of the feature points, so it is necessary to dynamically optimize the detection target area. After completing the hand tracking and recognition, 3D registration needs to be done by hand feature points. Extracting feature points include the following four parts: contour detection, feature point sampling, feature detection and feature point description.

A. Feature Detection

Operators commonly used in contour detection include: Roberts, Prewitt, Sobel[9], etc. Roberts operator uses adjacent pixel difference in the diagonal direction to detect the edge. The result is fine but is easily influenced by the noise. Prewitt is an edge model operator. It not only has good detection result, but also has a smooth effect on noise. Sobel uses the pixel weight algorithm in the up, down, left and right areas to finish grayscale detection. In this operator, noise can be further smoothed, while it needs much more calculation. Based on the practical situation in this paper, I decided to use Sobel algorithm to complete the hand contour detection.

The goal of feature point detection is to complete the three-dimensional registration. Feature point selection should satisfy the following two properties: detection convenience and scale invariance. There are many methods of feature point selection. During the natural interaction, since hands need to move or rotate, the feature point identification is easily affected of the environment. Corner detection doesn't have special requirements about the color of the detected image. No matter the color is rich or single, it can get good detection outcome. Besides, the calculation of corner detection algorithm is relatively smaller and the detection efficiency is higher. Commonly used algorithms include SUSAN, Plessey, Kitchen/ Resenfeld[10], curvature scale space algorithm and so on. Among these algorithms, in terms of curvature scale space corner detection algorithm, its calculation is the least, the detection speed is the fastest, and keeps high accuracy in the case of scale transformation. Therefore, this paper selects corner detection algorithm to detect hand feature points.

Since the mathematical formulas description of the hand contour curves can't be achieved, we need to use an approximate calculation of the curvature. In the reference, [11]Rosenfeld A and Johnston E proposed a method to describe the curvature at a single point on a curve approximately: giving three adjacent points on the curve, link these three points to construct triangle, calculate the angle

formed by the triangle at the middle point and use this angle to present the curvature of the curve here, which replaced the curvature of the curve approximately. As shown in a, in figure 1:

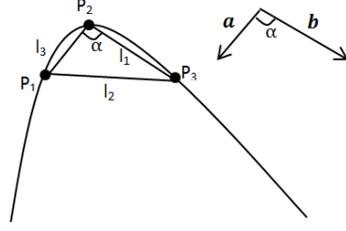


FIGURE I. CURVATURE CALCULATION

In the above figure, $P1(x_1,y_1)$, $P2(x_2,y_2)$, $P3(x_3,y_3)$ are the three consecutive sampling points. $P2, P3$ construct straight $I1$. $P1, P3$ construct straight $I2$. $P1, P2$ construct straight $I3$. Straight $I1$ and $I3$ build angle α . Define two vectors a and b , where a presents from $P2$ to $P1$ and b presents from $P2$ to $P3$. The vectors a and b are calculated by the following equations:

$$\mathbf{a} = (x_1 - x_2, y_1 - y_2) \quad (1)$$

$$\mathbf{b} = (x_3 - x_2, y_3 - y_2) \quad (2)$$

According to the cosine formula $\cos\alpha$:

$$\cos\alpha = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \quad (3)$$

Then the curvature at the point $P2$ can be approximately replaced by angle α :

$$\alpha = \arccos \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|} \quad \alpha \in [0, \pi] \quad (4)$$

After sampling a range of collection of feature points are obtained as shown in figure 2. In the set the curvature of each point of is sequentially calculated, and the feature point candidates which meet certain conditions are regarded as the final feature points of the curvature detection. They will be used in the three-dimensional registration calculation.

Through feature point detection, some areas of hands with a large curvature (with fingertips and finger point) will detect a number of feature points. These characteristic points are not invariant with rotation. It has big chance that detection missing or wrong detection happens when rotating the hand. It will influence the 3D registration later. In this paper, use the weighted average number of points to describe the general condition of the points in a region. After weighted average calculation, the robust of the description of feature points in the region improves, and the feature point description in different postures can be done accurately, greatly improving the adaptability of the hand.



FIGURE II. FINGERTIP DETECTION FEATURE POINT CANDIDATES

IV. DESCRIPTION OF FEATURE POINTS

As the fingertips and the root of fingers have very clear clustering features, so use data mining clustering algorithm to divide the feature points. Since the classification of the hand model in this paper is characterized to require that the same type of classification is relatively as compact as possible while the different types separate as much as possible. The models with such features commonly use K-MEANS[12] algorithm to classify.

K-MEANS algorithm is an iterative process. By clustering analysis, we can get a collection of hand clustering. Then, describe each cluster collection using the weighted average of the feature points. Through cluster analysis, we got a description including fingertip and root, total of nine feature points. As in the follow-3D registration our main object of study is the fingertip, so it is needed to exclude the feature points of the finger root points. At the last step, when we calculate the curvature of the feature points of hands we define two vectors. These two vectors of fingertip and finger direct differently (the peak located in the hand portion of the profile curve and bottom), so you can use the vector cross product pointing to different results (relative to the screen inside and outside) refers to the root point for exclusion.

Figure 3 is a shot of using K-MEANS algorithm to get fingertips candidate feature points clustering. In this figure the green dot is the fingertips points obtained from candidate fingertip points clustering.

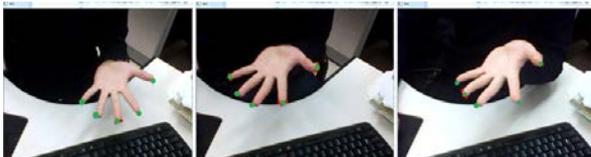


FIGURE III. CLUSTER ANALYSIS RESULTS

V. EXPERIMENT RESULT ANALYSIS

A. System Development Environment

The system is developed at 32-bit Windows PC platform. Logitech C270 HD webcam is used to video capture, of which the highest resolution is 1280x720.

System functions to be achieved are developed based on Microsoft Visual Studio 2008. In addition this paper uses the computer vision library OpenCV (Open Source Computer Vision Library), computer graphic rendering engine OpenGL (Open Graphic Library) and UG secondary development technology.

This paper improves CamShift adaptive mean shift tracking algorithm to obtain accurate color detection combined with two-dimensional image of the hand position. After using

Sobel operator edge detection to get the hand contour, through corner feature point sampling and extraction, K-MEANS poly class got five feature points fingertips and hands pose specific information, laying the foundation for the hand 3D registration in the natural interaction.

Experiment 1: In order to test the effectiveness and applicability of the improved process in this paper, an experiment is completed. It tested the accuracy and efficiency of the hand tracking identification. 5 participants were detected by tracking five different gestures for 20 times. Table 1 shows the detection rate of each gesture. Based on the operating results and experimental data in the table, the improved CamShift algorithm has good real and robustness, higher tracking speed. Detection window can completely cover the area of the hand. The detection rate is basically stable at around 78%.

TABLE I. THE DETECTION RATE OF HANDS TRACKING.

Gestures name	Total number of test	Just data
Palm up	100	83
Back up	100	77
Hand gestures 1	100	82
Hand gestures 2	100	74
Make a fist	100	75

Gestures name	Wrong data	Recognition rate
Palm up	17	83%
Back up	23	77%
Hand gestures 1	18	82%
Hand gestures 2	26	74%
Make a fist	25	75%

Experiment 2: In order to test this article to improve the effectiveness and applicability of the method, select 500 pictures of different actors used for the detection of different gestures as the test set to detect feature points in the image, compared to the traditional template matching, color testing and template matching, traditional CamShift add color detection. As shown in Table 2, it's known that the conventional template matching method is easy to implement, but because the scanning process exhaustive sampling and detection costs too much time, efficiency is not high, while the accuracy is not good; template matching combined with the detection of the color improves the detection rate and accuracy, but efficiency increases little; traditional CamShift combined with color detection has some advantages over the previous two methods in terms of efficiency, of which accuracy is good. In this paper, the efficiency of the improved method is similar

to conventional CamShift, especially the detection accuracy of a human finger portion has a raise, to get more accurate information on the position and orientation of hands in the complex context of enhanced assembly environment.

TABLE II. COMPARE AMONG PERFORMANCES OF DETECTION METHODS.

Method Name	Detection time(s)	Total number	Detected number
Template match	376	500	463
Complexion + Template match	349	500	457
Traditional CamShift	132	500	464
Improved CamShift	145	500	460

Method Name	Just data	Wrong data	Detection rate
Template match	374	89	74.8%
Complexion + Template match	382	75	76.4%
Traditional CamShift	431	33	86.2%
Improved CamShift	443	17	88.6%

Using the detected feature point information, combining with the camera intrinsic parameters model we have established the relationship between the real space coordinate projection system, the camera coordinate system and the image coordinate systems to obtain a relationship between the position and the hand relative position in the computer image in real space. So we can add in the appropriate location computer image with virtual 3D models of parts to complete the actual situation, play a guiding role in the assembly. Figure 4 is to use this method to get the hand feature points to guide the assembly, which is called augmented reality prototype system running legend.



FIGURE IV. A VIRTUAL OBJECT REGISTRATION DISPLAY

VI. CONCLUSION

This paper deeply studies the natural interaction of hands tracking and detection issue, especially about finger part ignored or undetected problems in the existing method. Propose an optimized adaptive mean shift algorithm combined with skin color detection to achieve accurate trace detection of hand. Considering the complexity of the background of environmental issues, this improved algorithm has better robustness and detection rate than the traditional methods, and it also obtains a more complete outline of the image of the hand, providing convenience for feature point extraction and

description later. It further improves the detection accuracy of the curvature of the hand from the extracted feature points. Using the K-Means clustering feature finger points, the feature points are effectively divided into groups. Use the weighted algorithm to complete the description of the feature point. Improve the rate of detection, ensuring that during the assemble process, a valid position and orientation information can be obtained during the continuous movement and rotation of the hand, which is inverted to complete registration of the subsequent three-dimensional.

REFERENCES

- [1] Sun Chao, Zhang Mingmin Hand Based Natural Interaction in Augmented Reality Journal of Computer-Aided Design & Computer Graphics Vol.23 No.4 Apr. 2011: 267-704;
- [2] Zhu Jiyu, Wang Xiyang, Wang Weixin, et al. Hand gesture recognition based on structure analysis [J]. Chinese Journal of Computers, 2006, 29(12): 2130-2137 (in Chinese)
- [3] Peng Tao Li Shiqi Wang Junfeng. Virtual Assembly Based on Augmented Human-Computer Interaction Technology. Journal of Computer-Aided Design & Computer Graphics [J], 2009, 21(3): 354-361,368.
- [4] Pan Z G, Li Y, Zhang M M, et al. A real-time multi-cue hand tracking algorithm based on computer vision [C]//Proceedings of IEEE Virtual Reality Conference. Los Alamitos: IEEE Computer Society Press, 2010: 219-223
- [5] Argyros A A, Lourakis M I A. Vision-based interpretation of hand gestures for remote control of a computer mouse [M]//Lecture Notes in Computer Science. Heidelberg: Springer, 2006, 3979: 40-51
- [6] Wang Xi-Ying, Dai Guo-Zhong. Recognition of Complex Dynamic Gesture Based on HMM-FNN Model. Journal of Software, 2008, 19(9): 2302-2312.
- [7] Liu YaZui Research of the augmented assembly, Master's thesis, Beijing university of posts and telecommunications 2013, 3.
- [8] Zuo Jun-Yi Ling Yan. Camshift Tracker Based on Multiple Color Distribution Models [J]. Acta Automatica Sinica, 2008, 34(7): [11]
- [9] Qiu-Qi Ruan, Digital image processing [M]. Electronic industry press, 2007.
- [10] Zhou Zhenhuan. Recognition of Object Shapes Based on Corner Features [J]. Computer Engineering, 2007, 33(06): 22-23,
- [11] ZHAO Wen-bin, ZHANG Yan-ning. Survey on Corner Detection [J]. Application Research of Computers, 2006, 23(10): 17-19.
- [12] Rosenfeld A, Johnston E. Angle detection on digital curves [J]. Computers, IEEE Transactions on, 1973, 100(9): 875-878.